

Linear multigrid techniques in self- consistent electronic structure calculations

Jean-Luc Fattebert

This article was submitted to
NATO Advanced Research Workshop on Multiscale computational
methods in chemistry and biology, Eilat, Israel, April 5-7 2000

U.S. Department of Energy

Lawrence
Livermore
National
Laboratory

May 23, 2000

DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This is a preprint of a paper intended for publication in a journal or proceedings. Since changes may be made before publication, this preprint is made available with the understanding that it will not be cited or reproduced without the permission of the author.

This report has been reproduced
directly from the best available copy.

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information
P.O. Box 62, Oak Ridge, TN 37831
Prices available from (423) 576-8401
<http://apollo.osti.gov/bridge/>

Available to the public from the
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Rd.,
Springfield, VA 22161
<http://www.ntis.gov/>

OR

Lawrence Livermore National Laboratory
Technical Information Department's Digital Library
<http://www.llnl.gov/tid/Library.html>

Linear multigrid techniques in self-consistent electronic structure calculations

Jean-Luc Fattebert

Center for Applied Scientific Computing (CASC)

Lawrence Livermore National Laboratory, Livermore, CA 94551, USA

E-mail: fattebert1@llnl.gov

Abstract. *Ab initio* DFT electronic structure calculations involve an iterative process to solve the Kohn-Sham equations for an Hamiltonian depending on the electronic density. We discretize these equations on a grid by finite differences. Trial eigenfunctions are improved at each step of the algorithm using multigrid techniques to efficiently reduce the error at all length scale, until self-consistency is achieved. In this paper we focus on an iterative eigensolver based on the idea of inexact inverse iteration, using multigrid as a preconditioner. We also discuss how this technique can be used for electrons described by general non-orthogonal wave functions, and how that leads to a linear scaling with the system size for the computational cost of the most expensive parts of the algorithm.

1 Introduction

The computational power of present supercomputers allows larger and larger calculations in quantum physics and chemistry. At the same time some applications require a very high accuracy, leading to new scientific computing challenges: computing hundreds of eigenfunctions of Hamiltonians discretized in subspaces of dimension of the order 10^6 . For example, an accurate description of a carbon nanotube composed of 160 atoms by the finite difference method requires 320 eigenfunctions on a grid composed of at least 250000 nodes.

Such calculations are actually possible thanks to the density functional theory (DFT) [1, 2, 3]. In DFT, the electrons are represented by independent orthonormal eigenfunctions ψ_j associated with the N lowest eigenvalues, $\epsilon_1 \leq \dots \leq \epsilon_N$, of the Kohn-Sham equations:

$$\left\{ \begin{array}{l} (-\frac{1}{2}\nabla^2 + \hat{V}[\rho])\psi_j(x) = \epsilon_j\psi_j(x), \\ \rho(x) = \sum_{j=1}^N f(\epsilon_j)|\psi_j(x)|^2, \\ \int_{\Omega} \psi_i(x)\psi_j(x)dx = \delta_{ij}, \\ \hat{V}[\rho] = \hat{V}^{ps} + V^H[\rho] + V^{xc}[\rho], \\ -\nabla^2 V^H = 4\pi\rho, \end{array} \right. \quad 1 \leq i, j \leq N, \quad (1)$$

in a domain Ω , with periodic or homogeneous Dirichlet boundary conditions. The function f defines the occupation of each orbital which can vary from 0 to 2. Here we neglect the spin. This allows us to deal only with half the number of electronic states.

The potential operator $\hat{V}[\rho]$ depends on the electronic wave functions through the electronic density ρ . It is composed of three terms. The first one, \hat{V}^{ps} , is the pseudopotential operator. It represents the effect of the atomic nuclei and core electrons on the valence electrons[4]. It means that only the valence electrons have to be described by the functions ψ_j , the other ones being included in the core. \hat{V}^{ps} is a sum of contributions from every atom in the system. For each atom α , the pseudopotential \hat{V}_α^{ps} is defined by its action on a wave function ψ ,

$$(\hat{V}_\alpha^{ps}\psi)(x) = v_\alpha(x)\psi(x) + \sum_{s=1}^{\bar{s}} \left(\int_{\Omega} \varphi_{\alpha,s}(x)\psi(x) \right) \varphi_{\alpha,s}(x).$$

Values for \bar{s} depend on the species of the atom (0 for Hydrogen, 1 for Carbon, 4 for Aluminium for example). The functions $\varphi_{\alpha,s}$ are radial and centered on the atoms. Though non-local, this operator preserves the sparsity of the Hamiltonian matrix in the calculation thanks to the short range of the functions $\varphi_{\alpha,s}$ ($\sim 3[bohr]$ for Carbon for instance). Note also that the functions v_α are smooth and — unlike the true Coulomb potentials — do not have singularities on the center of the atoms.

The second term, the Hartree potential $V^H[\rho]$, represents the Coulomb interaction between valence electrons. The last component of the potential, V^{xc} , describes — perfectly in principle — the exchange and correlation interaction between electrons. In practice, this term has to be approximated by a simplified model. Here we use the standard local density approximation (LDA)[5, 6].

A good way to discretize the Kohn-Sham equations is to describe the electronic orbitals and the potentials by their values on a uniform grid (number of nodes $\gg N$). The Laplacian operator is then approximated by finite differences (FD). Such an approach is particularly convenient for massively parallel implementations based on spatial domain decomposition [7]. It is also appropriate to implement local mesh refinements [8, 9, 10]. More recently, we have shown that it also allows to impose spatial localization constraints on the orbitals defined on the grid [11]. This essentially reduces the scaling of the calculation to $O(N)$ operations, the main goal of several algorithms developed recently (for recent reviews, see [12, 13]). Such an approach combines *ab initio* accuracy and scaling in an efficient way. Note that compared to the Finite Elements (FE) method, for a same accuracy, FD is in general cheaper on a uniform grid.

The fine mesh spacing used in practical calculations is very important to obtain accurate results. However, this represents a large number of degrees of freedom compared to the number of wave functions to compute. That also means a large condition number for the discretized Hamiltonian and thus requires an appropriate preconditioning to keep good convergence rates in the iterative eigensolver, with or without localization constraints on the orbitals.

After describing the basics of an inverse iteration method for the search of several eigenvalues in Section 2, we present in Section 3 a multigrid preconditioner that we use in an inexact inverse iteration for solving the KS equations. In Section 4, we discuss how such a linear preconditioner can be used along with non-orthogonal localized orbitals to obtain a linear scaling of the calculations.

2 Iterative eigensolver

The Kohn-Sham equations discretized by FD result in a huge 3D eigenvalue problem. Fortunately, the matrices involved are very sparse. Lanczos or Arnoldi based methods are very popular for large-scale eigenvalue problems with sparse matrices [14]. However

several features of DFT electronic structure calculations make these techniques not very suitable. First of all, a large number of eigenfunctions is in general required, making Krylov subspace expansions and deflation techniques very expensive. Secondly, for a self-consistent Hamiltonian, it is not useful to solve precisely the eigenvalue problem at each iteration. Cheap corrections in an appropriate direction are in general sufficient, and small variations of the Hamiltonian between self-consistent (SC) steps allow us to use the last computed wave functions as good initial approximations for the next step.

The Rayleigh Quotient Iteration (RQI) is a very fast algorithm to compute one eigenvalue of a matrix (see Ref. [15] for example). Here we describe this method in a different form that we find more suitable for matrices of size too large to make use of direct linear solvers. Starting from an approximate eigenpair $(\epsilon^{(k)}, \psi^{(k)})$ of a discretized Hamiltonian H , $\epsilon^{(k)}$ given by the Rayleigh Quotient of $\psi^{(k)}$ at step k , we look for a better approximation $\psi^{(k+1)}$. Specifically, we write

$$\psi^{(k+1)} = \frac{1}{\xi}(\psi^{(k)} + \delta\psi^{(k)})$$

where the correction $\delta\psi^{(k)}$ is chosen orthogonal to $\psi^{(k)}$. ξ is a normalization factor. Improving $\psi^{(k)}$ by RQI requires then to find $\delta\psi^{(k)} \perp \psi^{(k)}$ and ξ such that

$$(H - \epsilon^{(k)}) \frac{1}{\xi}(\psi^{(k)} + \delta\psi^{(k)}) = \psi^{(k)}.$$

We eliminate ξ by projecting the whole equation onto $(\psi^{(k)})^\perp$ and then rewrite it as

$$(I - \psi^{(k)}(\psi^{(k)})^T) (H - \epsilon^{(k)}) \delta\psi^{(k)} = - (H - \epsilon^{(k)}) \psi^{(k)}.$$

By the properties of the Rayleigh Quotient, the projector on the right hand side of the equation has been omitted. Note that in the Jacobi-Davidson method [16] the corrections are based on approximate solutions of the same equation.

This algorithm can be generalized for the simultaneous search of N eigenfunctions, with possible degeneracy of some eigenvalues [17, 18]. The idea is to replace the Rayleigh quotient by the Rayleigh-Ritz procedure [15] and to look for corrections orthogonal to the whole subspace $\Psi^{(k)}$ of trial eigenfunctions. Denoting $\Psi^{(k)} = (\psi_1^{(k)}, \dots, \psi_N^{(k)})$ as a matrix made of columns $\psi_j^{(k)}$, we write the iterative algorithm:

1. Let $\Psi^{(0)}$ be a trial subspace of dimension N .
2. For $k = 0, 1, 2, \dots$, do:
 - (a) Rayleigh-Ritz (RR) for H in the subspace $\Psi^{(k)} \rightarrow (\epsilon_j^{(k)}, \psi_j^{(k)}), j = 1, \dots, N$
 - (b) For $j = 1, \dots, N$, compute $\delta\psi_j^{(k)} \perp \Psi^{(k)}$ solution of

$$(I - \Psi^{(k)}\Psi^{(k)T})(H - \epsilon_j^{(k)})\delta\psi_j^{(k)} = -(H - \epsilon_j^{(k)})\psi_j^{(k)}, \quad (2)$$

- (c) Define $\Psi^{(k+1)} = \text{span}\{\psi_1^{(k)} + \delta\psi_1^{(k)}, \dots, \psi_N^{(k)} + \delta\psi_N^{(k)}\}$.

If we solve exactly Eq.(2), we obtain a locally quadratic convergence rate close to the solution of the eigenvalue problem for non self-consistent Hamiltonians [18]. For self-consistent calculations, we update the potential $V[\rho]$ at the end of the RR procedure by taking a mixing of the potential at step $k - 1$ with the one computed from the new density $\rho^{(k)}$.

3 Multigrid preconditioning

One can actually apply the algorithm of the previous section by solving the linear systems by multigrid [10]. But here we would like to show how one can replace the expensive solution of the linear system by a cheap and efficient preconditioner. The main advantage of this second approach is that it actually does not depend on j and works in any basis one can choose for the trial subspace $\Psi^{(k)}$.

Looking at the correction equation (2), we note that the right hand side is the steepest descent direction for the minimization problem with orthonormality constraints associated with the eigenvalue problem (1). We note it $(-\mathbf{r}_j^{(k)}, \mathbf{r}_j^{(k)})$ being the residual of the eigenvalue problem.

From the point of view of the inverse iteration method, the "optimal" correction is given by

$$\delta\psi_j^{(k)} = -((I - \Psi^{(k)}\Psi^{(k)T})(H - \epsilon_j^{(k)})\Big|_{\Psi^{(k)\perp}})^{-1}\mathbf{r}_j^{(k)}.$$

Thus we want to construct a preconditioner M which is a good approximation of the operator

$$(I - \Psi^{(k)}\Psi^{(k)T})(H - \epsilon_j^{(k)})\Big|_{\Psi^{(k)\perp}}.$$

If we have a close look at the Hamiltonian operator, we notice that for high energy states, the Laplacian is the dominant part, and the corresponding eigenfunctions are essentially similar to those of the Laplacian (Fig. 1) perturbed by a relatively weak potential. It means that $-\frac{1}{2}\nabla^2$ is a good approximation of $(H - \epsilon_j^{(k)})$ in $\Psi^{(k)\perp}$ for $1 \leq j \leq N$, so that one can choose

$$M \sim (I - \Psi^{(k)}\Psi^{(k)T})(-\frac{1}{2}\nabla^2)\Big|_{\Psi^{(k)\perp}}.$$

Preconditioners based on a similar idea have been developed for calculations in plane-wave bases [19, 20, 21]. Since the basis functions are then eigenfunctions of the Laplacian operator, it is easy to design such a preconditioner.

In real-space, one can associate frequencies with grid resolution. Applying a single grid iterative method — like Jacobi or Gauss-Seidel — to solve a Poisson problem, one essentially obtains the high frequency components of the solution, the one that we cannot represent on a coarser grid. Using multigrid V-cycles based on such a "smoother", we can solve the problem for components of lower frequencies by visiting coarser grids [22]. Furthermore, by choosing a limited number of grids, one can "select" the components that we want to solve for. Following this argument, we define the application of our preconditioner M^{-1} to $\mathbf{r}_j^{(k)}$ as an iterative multigrid solver for the Poisson problem

$$-\frac{1}{2}\nabla^2\delta\psi_j^{(k)} = -\mathbf{r}_j^{(k)}$$

limited to the grids with a number of nodes larger than $8N$. We start with an initial trial solution

$$\delta\psi_j^{(k)} = -\alpha I_{\ell-1}^\ell \dots I_{m+1}^{m+2} I_{m+2}^{m+1} \dots I_\ell^{\ell-1} \mathbf{r}_j^{(k)}. \quad (3)$$

Its main goal is to introduce some low frequency components in the correction. I_m^n denotes an interpolation from the grid m to the grid n [22]. We use the notation ℓ

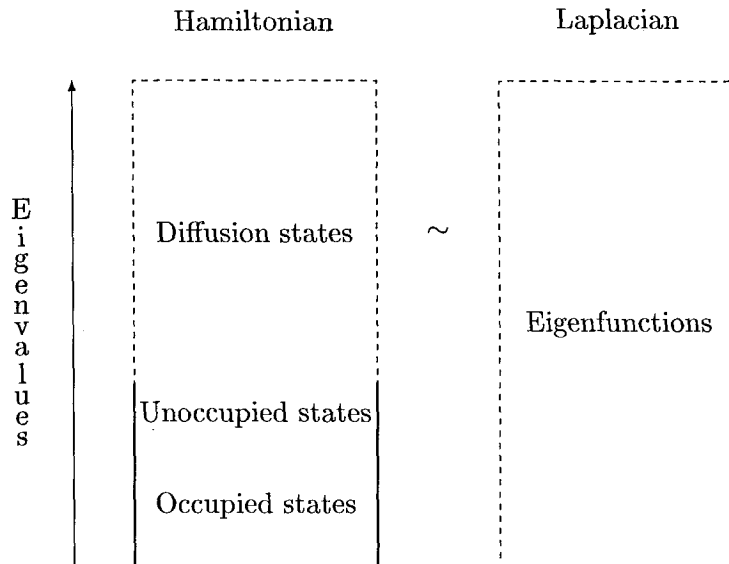


Figure 1: Preconditioning based on similarities between Hamiltonian and Laplacian operators for eigenfunctions associated with large eigenvalues.

for the finest grid and $m \leq \ell$ for the coarsest grid in the V-cycle. The coefficient α is defined by looking at Eq.(3) as the steepest descent correction one would make if the whole calculation was done on the grid $m - 1$, so that we choose $\alpha \sim h_{m-1}^2$. Here h_{m-1} is the mesh spacing on the grid $m - 1$. As a smoother we choose the Jacobi method because of its inherent parallelism.

Note that since $(-\mathbf{r}_j^{(k)})$ is the steepest descent direction,

$$\delta\psi_j^{(k)} = -M^{-1}\mathbf{r}_j^{(k)}$$

can be seen as a preconditioned steepest descent direction.

In Figure 2, we present the convergence history of the error on the energy for various discretization grids for an 8 atoms diamond cell self-consistent calculation. We use the multigrid preconditioner described above, doing for each correction $\delta\psi_j^{(k)}$ and at each SC iteration 1 V-cycle with 2 pre-smoothing and 2 post-smoothing. The grid-independence of the convergence rate is observed. All the calculations use the same coarsest grid $6 \times 6 \times 6$ for the multigrid preconditioning, and the same total number of states (occupied+unoccupied=24). The presence of unoccupied states is important to have a subspace $\Psi^{(k)\perp}$ in which the Laplacian is a good approximation of the Hamiltonian.

More numerical results and applications based on this method can be found in Ref.[11]. In the context of the Jacobi-Davidson method, a multigrid preconditioner has also been used to obtain grid-independent convergence rates on a simpler problem[23]. Note that in general, multigrid methods appear to be an essential component for large-scale electronic structure calculations in real-space (see Ref. [7, 10, 11, 24, 25]).

4 Linear scaling

The preconditioner described above is linear. It means that one can use it in a similar way for any basis of the subspace $\Psi^{(k)}$. For example in a basis of orthogonal functions

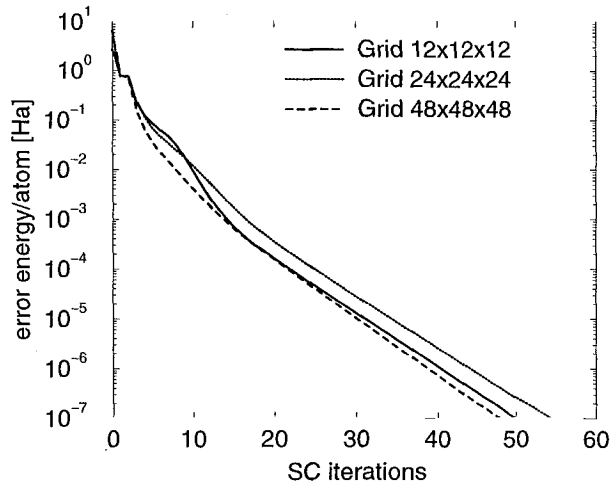


Figure 2: Convergence rate for a diamond cell calculation (8 atoms) for various discretization grids.

$\Phi = (\phi_1, \dots, \phi_N)$ related to Ψ by the transformation $\Phi U = \Psi$ where U is an $N \times N$ orthogonal matrix. Or for a basis of non-orthogonal functions $\chi = (\chi_1, \dots, \chi_N)$, where we have $\chi = \Phi G^T$. Here G is a lower triangular $N \times N$ matrix resulting from the Cholesky decomposition of $S = \Phi^T \Phi$. All these representations are mathematically equivalent since only the subspace spanned by these bases is important. Any algorithm expressed in a given basis can be simply expressed in another one by using the matrices U and G . We note that the evaluation of the electronic density is cheaper in the basis of the Ritz functions. With orthogonal functions, one also avoids problems related to a possibly ill-conditioned overlap matrix S . However, using non-orthogonal functions allows to impose localization constraints. Functions ϕ_i are allowed to be non-zero only inside spheres of radius R_c centered on the atoms for example. As shown in Ref.[11], if these constraints are not too strong — meaning if the localization radius of the orbitals is not too small — one obtains the same results one would get without localization within the desired accuracy. The computational cost for correcting the non-orthogonal functions, as well as for computing the electronic density ρ scales then linearly with the size of the system. There remains some $O(N^3)$ operations on $N \times N$ matrices, but not significant for calculations with N as large as 3000 [11].

The method described above can also be seen as a generalization of *ab initio* methods using LCAO (linear combination of atomic orbitals) or Gaussian basis functions:

$$\psi_j = \sum_i c_i \phi_i.$$

Here the local basis functions ϕ_i are defined by their values on a grid and are variationally optimized. These non-orthogonal orbitals described on a grid have much more degrees of freedom than atomic orbitals or Gaussians. Their accuracy can be systematically increased by mesh refinement or expansion of the localization domains. Since the number of orbitals per atom is somewhat minimal in this approach, the cost for the $O(N^3)$ part of the calculation remains small, at least for the size of the systems one can compute today. As shown in Ref.[11], this method allows accurate *ab initio* calculations with hundreds of atoms.

Acknowledgments

I would like to thank E. L. Briggs, J. Bernholc and F. Gygi for helpful discussions. A portion of this work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

References

- [1] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev. B*, 136:864–871, 1964.
- [2] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev. A*, 140:1133–1138, 1965.
- [3] R. M. Dreizler and E. K. U. Gross. *Density Functional Theory*. Springer, Berlin, 1990.
- [4] W. E. Pickett. Pseudopotential methods in condensed matter applications. *Comput. Phys. Rep.*, 9:115–198, 1989.
- [5] D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.*, 45:566–569, 1980.
- [6] J. Perdew and A. Zunger. Self-interaction correction to density-functional approximations for many electrons systems. *Phys. Rev. B*, 23:5048–5079, 1981.
- [7] E. L. Briggs, D. J. Sullivan, and J. Bernholc. Real-space multigrid-based approach to large-scale electronic structure calculations. *Phys. Rev. B*, 54(20):14362–14375, 1996.
- [8] F. Gygi. Electronic-structure calculations in adaptive coordinates. *Phys. Rev. B*, 48(16):11692, 1993.
- [9] N. A. Modine, G. Zumbach, and E. Kaxiras. Adaptive-coordinate real-space electronic-structure calculations for atoms, molecules, and solids. *Phys. Rev. B*, 55(16):1337–1340, 1997.
- [10] J.-L. Fattebert. Finite difference schemes and block Rayleigh Quotient Iteration for electronic structure calculations on composite grids. *J. Comput. Phys.*, 149:75–94, 1999.
- [11] J.-L. Fattebert and J. Bernholc. Towards grid-based $O(N)$ DFT methods: optimized non-orthogonal orbitals and multigrid acceleration. *Phys. Rev. B*, 2000. in press.
- [12] G. Galli. Linear scaling methods for electronic structure calculations and quantum molecular dynamics simulations. *Current Opinion in Solid State and Materials Science*, 1(6):864–874, 1996.
- [13] S. Goedecker. Linear scaling electronic structure methods. *Rev. Mod. Phys.*, 71(4):1085–1123, 1999.
- [14] Y. Saad. *Numerical Methods for Large Eigenvalue Problems*. Manchester University Press, Manchester, 1992.
- [15] B. N. Parlett. *The Symmetric Eigenvalue Problem*. Prentice-Hall, Englewood Cliffs, NJ, 1980.
- [16] G. L. G. Sleijpen and H. A. Van Der Vorst. A generalized Jacobi-Davidson iteration method for linear eigenvalue problem. *SIAM Matrix Anal. and Appl.*, 17(2):401–425, 1996.
- [17] J. Descloux, J.-L. Fattebert, and F. Gygi. RQI (Rayleigh Quotient Iteration), an old recipe for solving modern large scale eigenvalue problems. *Comput. Phys.*, 12(1):22–27, 1998.

- [18] J.-L. Fattebert. A block Rayleigh Quotient Iteration with local quadratic convergence. *ETNA*, 7:56–74, 1998.
- [19] M. P. Teter, M. C. Payne, and D. C. Allan. Solution of Schrödinger’s equation for large systems. *Phys. Rev. B*, 40(18):12255–12263, 1989.
- [20] G. W. Fernando, G.-X. Quian, M. Weinert, and J. W. Davenport. First-principles molecular dynamics for metals. *Phys. Rev. B*, 40(11):7985–7988, 1989.
- [21] N. Chetty, M. Weinert, T. S. Rahman, and J. W. Davenport. Vacancies and impurities in aluminium and magnesium. *Phys. Rev. B*, 52(9):6313–6326, 1995.
- [22] A. Brandt. Multilevel adaptative solutions to boundary-value problems. *Math. Comp.*, 31(138):333–390, 1977.
- [23] S. Oliveira. On the convergence rate of a preconditioned subspace eigensolver. *Computing*, 63:219–231, 1999.
- [24] F. Ancilotto, P. Blandin, and F. Toigo. Real-space full-multigrid study of the fragmentation of Li_{11}^+ clusters. *Phys. Rev. B*, 12(59):7868–7875, 1999.
- [25] J. Wang and T.L. Beck. Efficient real-space solution of the Kohn-Sham equations with multiscale techniques. *J. Chem. Phys.*, 112(21), 2000.